INSTITUTE FOR DEFENSE ANALYSES

IDA

Discrete Reliability Projection

J. A. Snyder, Project Leader
J. M. Fregeau
D. Spalding

IDA *The Institute for Defense Analyses is a non-profit corporation that operates three federally funded research and development centers to provide objective analyses of national security issues, particularly those requiring scientific and technical expertise, and conduct related research on other national challenges.*

INSTITUTE FOR DEFENSE ANALYSES

IDA Document D-4808

Discrete Reliability Projection

J. A. Snyder, Project Leader
J. M. Fregeau
D. Spalding

Executive Summary

This report discusses one method for calculating reliability growth projection for discrete systems. This method is simple enough to be applied by non-experts yet has significant advantages relative to the methods designed for continuous systems and applied to discrete systems. We summarize the assumptions and arguments that lead to the method, but do not delve into alternative assumptions that may be equally valid. We also compare and contrast this model with a well-known model for continuous systems.

Contents

1   Introduction

Reliability is defined as "the probability that an item will perform its intended function for a specified period of time..."(DoD 2011). As failure modes are discovered and corrected, the reliability can improve. This is called reliability growth: "the positive improvement in a reliability parameter over a period of time due to implementation of corrective actions to system design, operation or maintenance procedures, or the associated manufacturing process"(DoD 2011). If there is an active campaign to discover and correct failure modes, then it is possible to forecast reliability growth. This is called reliability growth projection: "an assessment of reliability that can be anticipated at some future point in the development program"(DoD 2011).

The reliability of systems is of great interest to manufacturers, users, and acquisition professionals. Since the true reliability is unknown, it is necessary to make an estimate of the reliability, that is, the best estimate of reliability at one point in time. This estimate typically consists of a point estimate and an associated confidence interval. The point estimate is equivalent to a single point on the reliability growth curve. Reliability estimates are usually made based on the results of one recent test event. Such tests are typically designed so that there are sufficient data to make a statistically significant reliability estimate (i.e., so that the associated confidence interval is not too large), although we note there is a tension with cost, since some tests can be expensive. For simplicity, the tests are typically designed with the assumption that the configuration of the item under test does not change during the test event.

If the system configuration changes during the test (for example, a fix is implemented), then the naive reliability estimate made from all test data will not properly reflect either configuration. In addition, the reliability estimates based on each portion of the test separately may not have the desired level of statistical significance. In cases such as these, reliability projection models can be used to make the best estimate of the system reliability at the end of the test.

Two components are key to reliability growth projection. The first is that "credit should be given" for fixed failure modes. In other words, if a failure mode is fixed after a test, the true reliability is greater than the naive estimate from the test, since a failure mode that appeared in that test has been fixed. The second is that the failure probabilities for the various failure modes in the system have a probability distribution ranging from very common to very rare. The failure modes that appear during the test are likely to be the most common (highest probability) failure modes, and there are likely to be some number of rarer failure modes that are not observed in the test. As common failure modes are fixed and additional tests are made, these rare failure modes may be observed, but many are likely to remain unobserved.

Reliability is usually measured against a continuous variable, such as time (e.g.,

mean time between failure [MTBF]). But for some systems, here called "one-shot" systems, reliability is measured against a discrete variable. Typical examples of one-shot systems are rockets, missiles, and torpedoes, where testing is done one firing at a time, and the length of time between shots is irrelevant.

This document discusses one implementation of discrete reliability growth projection and illustrates its use with examples.

2 Hall's Model

A recent Ph.D. thesis summarizes the history of the field of reliability growth and gives methodologies for specific aspects of discrete reliability growth (Hall 2008). For people new to the field, the first few chapters of this thesis are a valuable introduction to, and summary of, prior work. The later chapters are papers on reliability growth that were written for (self-contained) publication.

Of particular interest is Chapter 4, which describes a model for discrete reliability growth projection, published as Hall & Mosleh (2008). For continuous systems, reliability growth projection is a well-developed field and the models are well understood. There has been little work to date in discrete (one-shot) systems at the failure-mode level, however, and Hall's thesis appears to be a valuable contribution to the field.

The model assumes that a series of trials are undertaken where the outcome of each trial is either success or failure. Failures can occur for many reasons; these reasons are referred to as failure modes. Each failure mode has an underlying probability of appearance that is unknown. Failure modes are assumed to occur independently of each other. By construction, corrective actions (fixes) are not made in the middle of the series of trials—they are delayed until the end of the current test phase. Thus, the reliability of the system is constant during each test phase, and the fixes applied between test phases will cause the reliability to increase for each successive phase.

The data required to perform reliability projection are T, the total number of trials; N_i, the number of failures in the ith failure mode; and d_i, the fix effectiveness factor (FEF) for the ith failure mode. In practice, the FEFs are estimated (d_i^*), usually based on engineering judgement.

The following discussion summarizes the model developed in Chapter 4 of Hall (2008). We use the same notation and reproduce equations 12–30 from that chapter, although in a slightly different order.

The true failure probabilities (p_i) for each failure mode are unknown, but we can estimate the probability for each failure mode as:

$$\hat{p}_i = \frac{N_i}{T}. \tag{2.1}$$

The reliability point estimate $(\hat{r})$ at the end of the phase (before fixes) is simply the product of one minus these probabilities. If the probabilities are small then we can ignore second-order effects and approximate this as one minus the sum of the probabilities:

$$\hat{r} = \prod_i (1 - \hat{p}_i) \approx 1 - \sum_i \hat{p}_i = 1 - \frac{\sum_i N_i}{T}. \tag{2.2}$$

One of the difficulties with equation (2.2) is that it does not take into account the failure modes that exist but have not been observed in testing. For one-shot systems where testing is expensive, the number of trials is kept to a minimum, and rare failure modes are not likely to be observed. As a workaround for this, Hall's model requires that the user estimate k, the total number of failure modes inherent in the system. The true system reliability growth is therefore:

$$r(T) \equiv \prod_{i \in obs} \left[1 - (1 - d_i) \cdot p_i\right] \cdot \prod_{j \in obs'} (1 - p_j), \tag{2.3}$$

where obs is the set of all observed failure modes and obs' is the set of all unobserved failure modes. Note that for each observed mode where the FEF is 100% ($d_i = 1$), that mode contributes no unreliability. If all modes are observed and all FEFs are 100%, then the system reliability is 100%. Alternatively, if the FEFs are all zero (uncorrected) and all modes are observed, then the system reliability takes the same form as equation 2.2, the point reliability estimate.

Of course, the true failure probabilities (p_i) are unknown, and our estimates ($\hat{p}_i$) are only nonzero for observed failure modes. We must do something different for the unobserved modes. Hall used a shrinkage factor estimator given by:

$$\tilde{p}_i \equiv \theta \cdot \hat{p}_i + (1 - \theta) \cdot \frac{1}{k} \sum_{i=1}^{k} \hat{p}_i, \tag{2.4}$$

where θ is the (unknown) shrinkage factor, with $0 \le \theta \le 1$. Shrinkage is one method to transfer some of the measured unreliability from the observed modes to unobserved modes. There are alternate methods that should also work. Note that although Hall wrote the sum over the total number of failure modes (k), the sum effectively only runs over the observed failure modes since $\hat{p}_i = 0$ for all unobserved modes.

Hall assumes that the true failure probabilities come from a beta distribution[1] that can be parametrized as:

$$f(p_i) \equiv \frac{\Gamma(n)}{\Gamma(x) \cdot \Gamma(n - x)} \cdot p_i^{x-1} \cdot (1 - p_i)^{n-x-1}, \tag{2.5}$$

where n and x are the beta shape parameters, and Γ is the Euler gamma function. The parameters of the beta distribution can be estimated for our case using a moment-estimation technique (Martz & Waller 1991). Other methods could be used for this estimation, including maximum likelihood or Bayesian methods. The first and second sample moments are:

---

[1]The beta distribution is a family of continuous probability density functions whose shape is determined by two parameters. This is a reasonable—but not unique—choice for the underlying distribution of the true failure probabilities.

$$\bar{p}_u \equiv \frac{1}{k} \sum_{i=1}^{k} \hat{p}_i, \tag{2.6}$$

$$m_u^2 \equiv \frac{1}{k} \sum_{i=1}^{k} \hat{p}_i^2. \tag{2.7}$$

The estimates for the beta shape parameters (for a given value of k) are then:

$$\breve{n}_k = \frac{\bar{p}_u - m_u^2}{m_u^2 - \bar{p}_u/T - (1 - 1/T)\,\bar{p}_u^2}, \tag{2.8}$$

$$\breve{x}_k = \breve{n}_k \cdot \bar{p}_u. \tag{2.9}$$

The optimal shrinkage factor can be chosen by minimizing the expected sum of squared error between the estimated and true failure probabilities:

$$\frac{d}{d\theta} E\left[\sum_{i=1}^{k}(\tilde{p}_i - p_i)^2\right] = 0. \tag{2.10}$$

Hall reduces this to an expression that only depends on k and the mean and variance of the p_i, rather than on the p_i themselves. This reduces the number of unknowns from $k + 1$ to three. For the beta distribution (equation 2.5) the mean is x/n and the variance is $\frac{x \cdot (n-x)}{n^2 \cdot (n+1)}$. Thus, the shrinkage factor can be estimated (for a given value of k) as:

$$\breve{\theta}_k = \frac{1}{(\breve{n}_k/T)\,(1 - 1/k) + 1}. \tag{2.11}$$

Combining equations leads to a moment-based shrinkage factor estimate of p_i for finite k given by:

$$\breve{\tilde{p}}_{k,i} = \breve{\theta}_k \cdot \hat{p}_i + (1 - \breve{\theta}_k)\left(\frac{N}{k \cdot T}\right), \tag{2.12}$$

where $N \equiv \sum_i N_i$ is the total number of failures observed in the T trials.

This leads to a reliability growth projection estimate of:

$$\breve{\tilde{r}}_k(T) = \prod_{i \in obs} \left[1 - (1 - d_i^*) \cdot \breve{\tilde{p}}_{k,i}\right] \cdot \left[1 - (1 - \breve{\theta}_k) \cdot \left(\frac{N}{k \cdot T}\right)\right]^{k-m}, \tag{2.13}$$

where m is the number of observed failure modes and d_i^* estimates d_i (either based on a series of tests or by engineering judgment).

Since k is unknown, we look at the limiting behavior in complex systems where the number of failure modes approaches infinity. In that case the reliablity growth projection becomes

$$\breve{\breve{r}}_\infty(T) \equiv \lim_{k \to \infty} \breve{\breve{r}}_k(T) = \prod_{i \in obs} \left[1 - (1 - d_i^*) \cdot \breve{\breve{p}}_{\infty,i} \right] \cdot \exp\left[-(1 - \breve{\theta}_\infty) \cdot \left(\frac{N}{T} \right) \right], \quad (2.14)$$

where

$$\breve{n}_\infty \equiv \lim_{k \to \infty} \breve{n}_k = \frac{\sum_{i=1}^m \hat{p}_i - \sum_{i=1}^m \hat{p}_i^2}{\sum_{i=1}^m \hat{p}_i^2 - \sum_{i=1}^m (\hat{p}_i/T)}, \quad (2.15)$$

$$\breve{\theta}_\infty = \frac{T}{\breve{n}_\infty + T}, \quad \text{and} \quad \breve{\breve{p}}_{\infty,i} = \breve{\theta}_\infty \cdot \hat{p}_i. \quad (2.16)$$

We will use this later when discussing the sensitivity of reliability growth projection estimates to k.

3 Examples

3.1 Example 1

In Chapter 4 of his thesis, Hall (2008) provides an example of discrete reliability growth projection that we will repeat here. He notes that it comes from an unspecified air-to-ground missile program, but the origin is not important. The data are as follows. There were $T = 27$ trials resulting in 11 failures that were ascribed to $m = 7$ different failure modes. The distribution of failures among the modes is given in Table 3.1, along with the estimated FEF for the corrective actions taken after the testing.

Table 3.1: Input Data (Failure Counts and FEF Estimates) for Example 1 (Hall 2008)

Mode	Failures N_i	FEF d_i^*
1	1	0.95
2	1	0.70
3	1	0.90
4	1	0.90
5	4	0.95
6	2	0.70
7	1	0.80

Using equations 2.1 and 2.2 we can calculate the failure mode probabilities for each mode and the reliability point estimate. For the unobserved modes, $\hat{p}_i = 0$, and they do not contribute to the measured unreliability of the system. Following Hall we choose the total number of failure modes to be $k = 50$. This allows us to calulate the moments, shape parameters, and shrinkage factor that lead to the best estimate for the probabilities, $\breve{\tilde{p}}_{k,i}$. Some of these quantities are summarized in Table 3.2.

$$\bar{p}_u = 0.008148 \qquad m_u^2 = 0.0006859. \tag{3.1}$$

$$\breve{n}_k = 23.3084 \qquad \breve{x}_k = 0.1899. \tag{3.2}$$

$$\breve{\theta}_k = 0.5417. \tag{3.3}$$

From these we can now calculate the best estimate of the reliability growth projection:

Table 3.2: Input Data and Calculated Intermediate Quantities for Example 1

Mode	Failures N_i	Meas. Prob. $\hat{p}_i$	Est. Prob. $\breve{\tilde{p}}_{k,i}$	FEF d_i^*	Mode Reliability $1 - (1 - d_i^*)\breve{\tilde{p}}_{k,i}$
1	1	0.0370	0.0238	0.95	0.9988
2	1	0.0370	0.0238	0.70	0.9929
3	1	0.0370	0.0238	0.90	0.9976
4	1	0.0370	0.0238	0.90	0.9976
5	4	0.1481	0.0840	0.95	0.9958
6	2	0.0741	0.0439	0.70	0.9868
7	1	0.0370	0.0238	0.80	0.9952

$$\breve{\tilde{r}}_k(T) = [0.9653] \cdot [0.9963]^{43} = [0.9653] \cdot [0.8514] = 0.8218. \tag{3.4}$$

As expected, this projection (which includes fixes) is significantly higher than the naive reliability point estimate (which does not contain fixes) of $16/27 = 0.5926$.

3.2 Example 2

Here, we provide an example of how the reliability growth projection can be used to reduce statistical uncertainty by combining trials that otherwise could not be combined. In this case we have an unspecified missile system where a testing phase was sized to obtain a specific statistical uncertainty in the reliability. However, during the testing a significant failure mode was discovered and diagnosed, and a fix was implemented in the middle of the test. Since the configuration of the device under test has changed, it is no longer legitimate to combine all the trials and calculate the reliability naively based on the full set of trials. Instead, each portion of the test must be treated separately, which will increase the uncertainty on the reliability for each portion to an unsatisfactory level.

Part A consisted of $T = 11$ trials with two failure modes discovered, one of which was fixed (Table 3.3). Again, we assume $k = 50$ total failure modes.

Table 3.3: Input Data and Calculated Intermediate Quantities for Example 2, Part A

Mode	Failures N_i	Meas. Prob. $\hat{p}_i$	Est. Prob. $\breve{\tilde{p}}_{k,i}$	FEF d_i^*	Mode Reliability $1 - (1 - d_i^*)\breve{\tilde{p}}_{k,i}$
1	2	0.1818	0.08024	1.00	1.0000
2	1	0.0909	0.04169	0	0.9583

$$\bar{p}_u = 0.005454 \qquad m_u^2 = 0.0008264. \tag{3.5}$$

$$\breve{n}_k = 15.2471 \qquad \breve{x}_k = 0.08316. \tag{3.6}$$

$$\breve{\theta}_k = 0.4240. \tag{3.7}$$

$$\breve{\breve{r}}_k(T) = [0.9583] \cdot [0.9969]^{48} = 0.8240 \qquad \text{part A only.} \tag{3.8}$$

Part B consisted of $T = 13$ trials with two failure modes discovered, neither of which was fixed (Table 3.4). We still assume $k = 50$ total failure modes.

Table 3.4: Input Data and Calculated Intermediate Quantities for Example 2, Part B

Mode	Failures N_i	Meas. Prob. $\hat{p}_i$	Est. Prob. $\breve{p}_{k,i}$	FEF d_i^*	Mode Reliability $1 - (1 - d_i^*)\breve{p}_{k,i}$
1	1	0.0769	0.03478	0	0.9652
2	2	0.1538	0.06686	0	0.9331

$$\bar{p}_u = 0.004615 \qquad m_u^2 = 0.0005917. \tag{3.9}$$

$$\breve{n}_k = 18.5375 \qquad \breve{x}_k = 0.08555. \tag{3.10}$$

$$\breve{\theta}_k = 0.4171. \tag{3.11}$$

$$\breve{\breve{r}}_k(T) = [0.9007] \cdot [0.9973]^{48} = 0.7914 \qquad \text{part B only.} \tag{3.12}$$

For part A, the reliability point estimate is $r = 8/11 = 0.7273$. The associated confidence interval can be calculated using standard binomial statistics.[2] For part A the 80% confidence interval is $[0.4892, 0.8952]$. For part B the point estimate is $r = 10/13 = 0.7692$ with an 80% confidence interval of $[0.5557, 0.9120]$. The reliability growth projections from each of part A and part B are larger than these naive point estimates. To combine the two reliability growth projections into a single projection with an associated confidence interval, one can perform the following thought experiment. If part A were performed with the failure mode fix in place, we would expect 9.06 successful trials instead of 8. We could then combine parts A and B, giving us an overall reliability point estimate of $r = (9.06 + 10)/(11 + 13) = 19.06/24 = 0.7942$ with an 80% confidence interval (determined using standard binomial statistics) of $[0.6503, 0.8969]$. This statistical uncertainty is significantly smaller than the uncertainty associated with either reliability growth projection individually.

---

[2]We use the MATLAB function `binofit`, which uses the Clopper-Pearson method.

3.3 Sensitivity to k

For the examples above we chose to set the total number of failure modes to $k = 50$. This was an arbitrary choice based on the fact that there were only a small number of failure modes observed, yet both examples were of complicated systems where there could potentially be a large number of failure modes. It is reasonable to ask how sensitive the reliability growth projection is to our choice of k. For example 1 we find that once k is substantially larger than m (the number of observed failure modes),[3] the reliability growth projection is fairly insensitive to the actual value of k (see Table 3.5 and Figure 3.1 for details). The difference between $k = 100$ and $k = \infty$ is only 0.4% and between $k = 50$ and $k = \infty$ is only 0.8%. Example 2 shows the same behavior.

Table 3.5: Sensitivity of the Reliability Growth Projection to k for Example 1

k	$\breve{\tilde{r}}_k(T)$
7	0.9348
15	0.8413
30	0.8260
50	0.8218
100	0.8185
1000	0.8158
10000	0.8155
∞	0.8155

3.4 MATLAB Code

We have written a simple MATLAB (The MathWorks, Inc. 2013) script that calculates the reliability growth projection for a fixed set of data from a single test series. All one needs to do is change `m`, `k`, `T`, `Ni`, and `di` and run the script. The result, `rk`, is given by equation 2.13. The code is on the CD accompanying this document.

```
%  Reliability growth projection based on Hall's thesis, Chapter 4
%
%  Numbers for Example 1
m = 7;   % Number of observed failure modes
k = 50;  % Total number of failure modes
T = 27;  % Number of trials
Ni = [1 1 1 1 4 2 1];  % Number of failures per mode
di = [0.95 0.70 0.90 0.90 0.95 0.70 0.80];  % Estimated FEF per mode
%
```
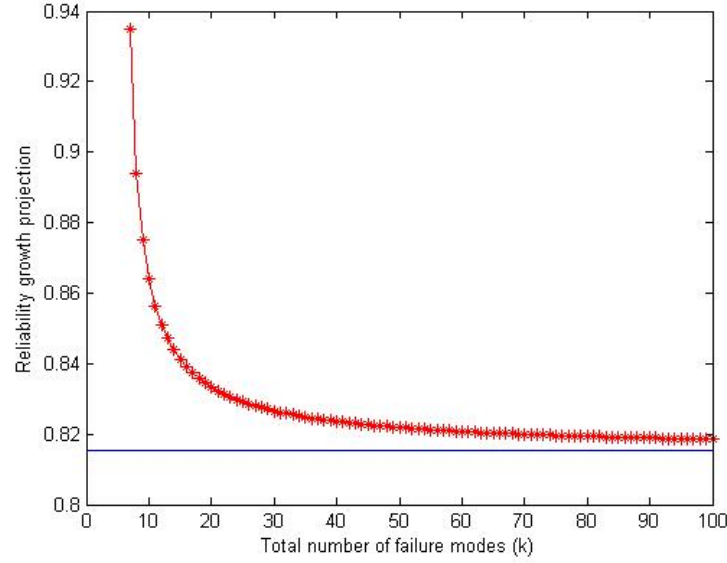
---

[3]Hall suggests $k > 5m$.

Figure 3.1: The Reliability Growth Projection for Example 1 as a Function of the Total Number of Failure Modes (k). The value for $k = \infty$ is plotted as the blue horizontal line.

```
N = sum(Ni);  % Total number of failures
pi = Ni/T;    % MLE of probability of failure per mode
pu = (1/k)*sum(pi);
mu2 = (1/k)*sum(pi.*pi);
nk = (pu-mu2) / (mu2 - pu/T - (1-1/T)*pu^2);
xk = nk*pu;
thetak = 1/( (nk/T)*(1-1/k) + 1);
pki = thetak*pi + (1-thetak)*(N/(k*T));
rki = 1 - (1-di).*pki;
r1 = prod(rki);
r2 = (1 - (1-thetak)*(N/(k*T)))^(k-m);
rk = r1*r2;
fprintf('Reliability growth projection is %f\n', rk);
```

3.5   Excel Code

We have put the same set of equations into a simple Excel (Microsoft Corporation 2013) spreadsheet. All one needs to do is change the m, k, and T values in A1, A2, and A3, then press "REFORMAT." This will create m lines for the N_i and d_i, starting at B6 and C6. The results are automatically calculated. The spreadsheet is on the CD accompanying this document.

4 Comparing and Contrasting Hall Discrete and ACPM Continuous Models

Section 2 develops the mathematical foundation of the Hall discrete model (HDM). The model's premise is that test data are binomially distributed and are from a finite number of discrete pass/fail trials. It is not clear, however, if the model remains valid if the number of trials increases arbitrarily (i.e., the test variable becomes continuous rather than discrete). More generally, what are the factors that determine the HDM's scope of applicability?

In the following, this issue is explored by comparing and contrasting the HDM with a continuous model, the AMSAA-Crow Projection Model (ACPM). For the two models, Table 4.1 lists fundamental features. Each model addresses two issues for reliability projection—accounting for unsurfaced failure modes and representing the statistical distribution of failure rates.

Table 4.1: Fundamental Features of the HDM Discrete and ACPM Continuous Models

Factor	HDM	ACPM
Test Type	Discrete	Continuous
Corrective Action	End of Test	End of Test
Statistical Mode	Binomial	Poisson
Unsurfaced Failure Modes	Shrinkage Factor	Fit to Data
Failure Likelihood	β distribution	Power Law
Test Inputs	Number of Trials; Number of Failures; Fix Effectiveness Factor	Length of Test; Number of Failures; Time of First Occurrence; Fix Effectiveness Factor
Output	Reliability	Mean Time Between Failure

The models' outputs are expressed differently, but can be inter-converted by (1) matching test variables and (2) translating between the HDM discrete probability of success per trial to the (quasi) equivalent ACPM mean time between failure (MTBF). The independent variables are matched by assigning a given length of time t (or other continuous measure of test duration) to each HDM trial. For n trials, the equivalent continuous test is $n \times t$ hours. The probability of success for a single trial p is translated into MTBF by compounding p in n consecutive trials. The result is equivalent to the probability of zero failures for the continuous test of length $n \times t$. The inter-conversion is completed by using the probability of zero failures to calculate MTBF. Assuming the Poisson distribution for failures in the continuous case, the whole process is expressed by the formula $\text{MTBF} = -t/\ln(p)$.

Table 4.2 summarizes features of the models' algorithms. HDM is deterministic and has no explicit features for testing the statistical validity of its results. However, it is amenable to sensitivity analyses using variations of its inputs. ACPM is inherently statistical and provides explicit tests for goodness of fit to the data, but the statistical power of the tests depends on the quantity of available data. Also, ACPM includes the time of first occurrence for each surfaced failure mode. HDM does not use information on where in a sequence of trials the failures occur. With respect to fix effectiveness factors, HDM uses the specific value for each failure mode while ACPM averages the input values prior to use in its algorithm.

Table 4.2: Algorithmic Aspects of HDM and ACPM

Algorithm Factor	HDM	ACPM
Treatment of Uncertainty	Deterministic	Statistical Fit to Data
Goodness of Fit Testing	No	Chi Square and Cramer-von Mises
Time of Failure Occurrence	Not Included	Included
Treatment of Fix Effectiveness	Uses Individual Values	Averages Individual Values

4.1   HDM and ACPM Characteristics

The following discussion addresses statistical models, inclusion of unsurfaced failure modes, and failure parameter likelihood functions.

4.1.1   Statistical Models

Both projection models assume that the system under test has a number (k) of independent failure modes that are either manifested ("surfaced") or unexpressed in a test. Their statistical models reflect whether the independent test variable is discrete or continuous.

The **HDM** assumes that the data are are from a finite series of discrete trials and that failures from individual failure modes are statistically independent and binomially distributed. The probability of success (reliability) after corrective actions is the product of the probability of non-occurrence for all failure modes (equation 2.3).

The **ACPM** assumes that the data are from a single test whose length is measured by a continuous independent variable such as time or miles. Failures for each mode are statistically independent and are individually Poisson distributed with constant failure rates whose initial values are λ_i. Total initial failure rate $[\rho(0)]$ is the sum of the individual initial failure rates:

$$\rho(0) = \sum_{i=1}^{k} \lambda_i \tag{4.1}$$

After corrective actions at time (or other continuous test variable) T, with FEF μ_i for each failure mode, that sum becomes:

$$\rho(T) = \sum_{i=1}^{k} (1 - \mu_i)\lambda_i \qquad (4.2)$$

The underlying assumptions of both statistical models are not inconsistent, the principal difference being the mathematical consequence of discreteness vice continuity of the independent variable. As is generally known, in the limit of an infinite number of trials n with binomial probability $p \longrightarrow 0$ such that $p * n$ is constant $= \lambda$, the binomial distribution approximates the Poisson distribution with parameter λ.

4.1.2   Unsurfaced Failure Modes

Given these assumptions, the statistical models for HDM and ACPM are valid. However, a problem arises when test data are directly used with the projection models.

For the **HDM**, in a test, unobserved failure modes have no failures to be used for estimating their failure probabilities. The most direct estimate for these probabilities is zero. But if fix effectiveness factors are used, this would lead to a biased estimate of reliability. An easy way to see this is to consider the limiting case where all fix effectiveness factors are 1.0 and all unsurfaced modes are estimated to have zero probability of failure. In that case, equation 2.3 shows that reliability is 100 percent. Thus, because there are generally unsurfaced failure modes that decrease true reliability below 100 percent, a policy where surfaced failure modes are completely fixed and unsurfaced modes are assumed to have zero failure rates will generally overestimate system reliability. When this occurs, regardless of how many times a test is repeated, the pattern of estimated reliabilities will not average out to the true reliability. This is the essence of a biased estimator. Thus, an unbiased estimator needs to account for the effect of unsurfaced failure modes. Equation 2.4 accounts for unobserved modes by adjusting the estimate of probability for all modes. The probability of failure of each mode (left side of 2.4) becomes a weighted average of (1) the observed failure rate for that mode (first term on the right) and (2) the average of rates for all modes (second term on the right). The weighting factor, which is in the interval [0,1], is called the shrinkage factor and is determined using an error-minimizing procedure (equations 2.6–2.11). This causes unsurfaced modes to appear in the reliability estimate because their adjusted values are non-zero despite exhibiting no failures during a test.

Equation 2.4 also changes the estimated reliability of each surfaced failure mode, with the first term on the right decreasing the observed probability of failure but the second term increasing it by adding a fraction of the average failure probability. An interesting property of the combination of terms in equation 2.4 is that the statistical expectation of the sum of the adjusted probabilities of failure equals the sum of the true probabilities of failure. That is, that sum is an unbiased estimator of the sum of the probabilities of failure, and by equation 2.2, using the adjusted reliabilities gives an unbiased estimate of system reliability prior to corrective action.

The **ACPM**[4] divides failure modes into those that are not intended to be fixed (A modes) and those that will be fixed after a test, if they surface in the test (B modes). Thus, a modified expression for the statistical model is:

$$\rho(T) = \lambda_A + \sum_{i=1}^{B}(1 - \mu_i)\lambda_i \tag{4.3}$$

where the second term is a sum over all B modes and $A + B = k$.

A direct way to apply test data to equation 4.3 is to estimate the individual variables in the equation. Engineering judgment can estimate the μ_i as numbers d_i, and the failure rates are estimated as $\hat{\lambda}_i$ equal to the number of failures for each B mode divided by the total test length. If this is done, then the estimated total failure rate after corrective actions is:

$$\hat{\rho}(T) = \lambda_A + \sum_{i=1}^{B}(1 - d_i)\hat{\lambda}_i \tag{4.4}$$

While plausible, the above is a biased estimate of $\rho(T)$. The expectation of $\hat{\rho}(T)$ is:

$$E\left[\hat{\rho}(T)\right] = \lambda_A + \sum_{i=1}^{B}(1 - d_i)\lambda_i + \sum_{i=1}^{B}d_i\lambda_i e^{-\lambda_i T} \tag{4.5}$$

The third term on the right side of equation 4.5 is the correction for bias in equation 4.4. Based on a sequence of steps (Crow 1983), the derivation of ACPM estimates that term as:

$$\sum_{i=1}^{B}d_i\lambda_i e^{-\lambda_i T} = \mu_d \lambda \beta T^{\beta-1} \tag{4.6}$$

where μ_d is the average fix effectiveness factor, and λ and β are parameters of the failure likelihood power law that are determined by a fit to test data.

Thus, HDM and ACPM have alternative ways of including unobserved failure modes and therefore addressing statistical bias. It appears neither has clear advantages, although the ACPM estimate seems to have a stronger case for an unbiased estimate of post-fix reliability. Note that the HDM correction for unobserved failure modes has no restriction on the number of trials, and it appears to apply to the quasi-continuous case of a large number of trials.

---

[4]The equations in this section are adapted from Crow (1983).

4.1.3   Failure Likelihood

To account for unsurfaced failure modes, HDM and ACPM assume a likelihood function for the probability parameters in their statistical models.

In the **HDM**, the two parameters of the beta likelihood of failure probabilities (equation 2.5) are estimated from the first two moments of the test data (equations 2.6–2.9). The example in Section 3.1 provides data to illustrate implications of using this likelihood function. The example has 27 trials, 7 B-modes, and 11 failures for those modes. Using the example's parameters, Figure 4.1 shows the cumulative likelihood, the likelihood density, a predicted likelihood, and the empirical likelihood distribution for test data.
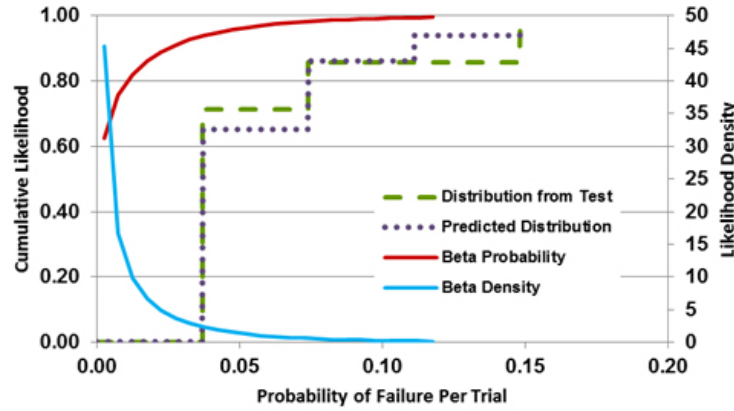

Figure 4.1: Likelihood for Example 1 (Section 3.1)

The horizontal axis in the figure shows values of probability of failure per trial, and the vertical axis shows the likelihood for a failure mode to have the indicated value.

The upper continuous curve (red) is the beta distribution that describes the cumulative likelihood of failure probabilities for the example in Section 3.1. Note that the most likely failure probabilities are small. Roughly 90 percent of the failure probabilities are less than or equal to 0.037, the smallest value that can be observed in 27 trials (1 failure/27 trials). The lower continuous curve (light blue) is the beta failure density. The right-hand vertical scale shows its values. All but a small fraction of the density is for failure probabilities less than or equal to 0.037.

The dotted line (dark blue) is the result of numerical integration of beta and binomial distributions. It shows a theoretical prediction of the outcome of the 27 trials in the example, given the beta distribution in failure probability and the fact that only 1 or more integer numbers of failures are observable in 27 trials. (It is also a fit because the parameters of the beta distribution are derived from the data.) The line is merely a guide for the eye—each vertical step represents a value in a distribution of discrete values. For example, the first vertical line indicates that about 65 percent of the observed failure modes should exhibit only 1 failure in 27 trials and have estimated

probability of failure equal to 0.037. The second step, for failure modes that exhibit 2 failures in 27 trials, adds approximately 20 percent to the cumulative distribution, and so on.

The dashed line (green) is the distribution of observed probabilities in the example. It agrees well with the theoretical prediction, indicating that use of the beta distribution and the method of moments is reasonable.

The **ACPM** estimates the correction term for unsurfaced modes, the rightmost sum in equation 4.5, by the power law in equation 4.6. This implicitly performs the summation in 4.5, thus implicitly assuming a likelihood of the λ_i. Because the estimation procedure reduces free parameters to only two (λ and β) , there may be no unique mapping of the power law into a likelihood function. With ACPM there is therefore no direct way to compare the empirical likelihood of failure probabilities with the assumed theoretical one. ACPM includes two statistical tests of fit for its model, however, which is an implicit way of checking the validity of the underlying likelihood function.

4.1.4 Other Factors

For the **HDM**, equations 2.6, 2.7, 2.11, 2.12 and 2.13 depend on k, the total number of failure modes. The model user needs to specify its appropriate values. The default is a value that is great enough for HDM to approach its asymptote for estimated reliability.

As currently formulated, HDM cannot be used unless at least one surfaced failure mode has more than one failure. Otherwise, all surfaced modes have an estimated probability of 1/T, leading to a negative value for n_k, the number of pseudo-trials in the beta likelihood for failure probability. The beta likelihood is undefined for negative numbers of pseudo trials.

For the **ACPM**, the estimation of reliability requires a fit to test data. While the end result is a reliability estimate for the corrected system at the end of test, a by-product is an estimate of the power-law parameters that can support a prediction of the number of surfaced modes as a function of time (or other independent variable).

4.2 Example Model Applications

For the example in Section 3.1, Table 4.3 compares HDM and ACPM estimates of reliability. The HDM estimate of reliability is from that section. For its estimate, ACPM is used assuming the 27 trials equate to 27 hours and that the 11 failures are randomly distributed in that interval. In the table, the values for ACPM include the median values for 100 random trials, with an 80-percent confidence interval in parentheses.

The table indicates that for this case, the models produce substantially differing estimates of MTBF, but comparatively close estimates of single-trial reliability. The

Table 4.3: Comparison of Projection Model Predictions for Example 1 in Section 3.1

Projection Model		Reliability	MTBF
HDM		0.8218	5.095
ACPM	Using $\hat{\beta}$	0.745 (0.652-0.803)	3.39 (2.34-4.55)
	Using $\bar{\beta}$	0.769 (0.686-0.821)	3.82 (2.65-5.06)

reason for this is that reliability cannot exceed 100 percent but MTBF can increase indefinitely. Thus, small differences in single-trial reliability can equate to large differences in MTBF. ACPM has alternative estimates of MTBF, depending on whether the estimate is based on a maximum likelihood parameter $\hat{\beta}$ or an unbiased $\bar{\beta}$. Agreement between models is better using $\bar{\beta}$. The HDM reliability and MTBF estimates fall at the upper end of ACPM 80-percent confidence intervals.

Table 4.4 extends the comparison of HDM and ACPM to a range of input values for B modes (m), total failures (N), and trials (T). The 21 lines in the table are in 3 groups of 7 cases each. Each group has numbers of trials that vary from 5 to 100 but a fixed number of failures. Thus, for a constant total test in hours, all cases in a group should have approximately the same overall ratio of failures to time (i.e., MTBF). The output that changes within each group of seven is the single-shot reliability—for a constant number of failures, as the number of trials increase, the failure rate per trial decreases, and consequently the reliability per trial increases.

For each case, the table lists single-shot reliability from the **HDM**, an equivalent HDM MTBF, and a comparable ACPM MTBF. The HDM MTBF is derived assuming a fictional discrete system that is reused over its m trials and ends with a probability of successful completion of all trials. Assuming a 100-hour total test time and an exponential probability of successful completion, this translates into a hypothetical MTBF for all trials (i.e., for the entire 100 hours). The table lists ACPM results based on the unbiased estimate of β.

In the table, most of the cases in each group of seven have almost the same MTBF, and that value agrees with the ACPM median MTBF to between a few and about 10 percent. Intuitively, the MTBF difference between HDM and ACPM should be least with the greatest number of trials (100). This is because the greater the number of trials, the more closely the binomial distribution assumed for the HDM approximates the Poisson distribution assumed for the ACPM. However, for the trial data in Table 4.4, this trend in results does not always occur. One relevant factor appears to be the sensitivity of ACPM MTBF estimates to the distribution of failure mode first occurence times. Because the HDM model does not require such times, it is not clear which are the appropriate first occurence times for comparison. Also, other choices for trial data might have been more suited to the power law in equation 4.6. Nonetheless, within the scope of this analysis, the HDM and ACPM results do not disagree. In each of the three cases in Table 4.4, the HDM MTBF values are well within the 80-percent

Table 4.4: Comparison of HDM and ACPM Results For Ranges of Input Parameters

HDM Quantities			HDM Reliability	HDM MTBF	ACPM MTBF
B-modes	Failures	Trials	Single Shot Reliability After Corrective Action	MTBF for 100 Hour Continuous Test (Hypothetical)	Using $\bar{\beta}$ (Median/80% Sample Interval)
3	5	5	0.540	32.5	33.2 (22.3-43.7)
		10	0.720	30.5	
		15	0.801	30.0	
		25	0.874	29.6	
		50	0.936	30.2	
		75	0.957	30.1	
		100	0.967	30.1	
6	10	5	0.292	16.2	16.3 (10.9-20.6)
		10	0.519	15.2	
		15	0.641	15.0	
		25	0.763	14.8	
		50	0.872	14.7	
		75	0.925	14.6	
		100	0.934	14.6	
9	15	5	0.157	10.8	9.68 (6.87-11.5)
		10	0.374	10.2	
		15	0.513	10.0	
		25	0.667	9.9	
		50	0.815	9.8	
		75	0.872	9.7	
		100	0.902	9.7	

confidence intervals for the ACPM MTBF sample.

4.3   Observations

Despite quite different initial assumptions, reliability projections from HDM and ACPM are similar. It appears the choice between these models would depend on how well basic assumptions apply. HDM is strongest in its use of fix effectiveness factors and would seem to be best for small numbers of trials, especially if there are few surfaced failure modes. In such cases results could be sensitive to how the factors are distributed among the modes. ACPM seems best when there is sufficient testing to uncover many failure modes, thus making an average fix effectiveness factor more representative of a central tendency and also providing enough time-of-occurrence data to facilitate a reasonably precise estimate of failure rate vs. test time.

References

Crow, L. H., "Reliability Growth Projection from Delayed Fixes," Proceeding of the Annual Reliability and Maintainability Symposium, pp. 84–89, 1983

DoD, "Reliability Growth Management," U.S. Department of Defense, Handbook MIL-HDBK-189C, 2011

Hall, J. B., *Methodology for Evaluating Reliability Growth Programs of Discrete Systems*, Ph.D. thesis, University of Maryland, College Park, 2008

Hall, J. B. & Mosleh, A., "A Reliability Growth Projection Model for One-Shot Systems," IEEE Transactions on Reliability, Vol. 57, pp. 174–181, 2008

Martz, H. F. & Waller, R. A., "Selecting Beta Prior Distributions based on Observed Binomial Sampling Data," in *Bayesian Reliability Analysis*, 2nd ed., Malabar, FL: Krieger, 1991

Microsoft Corporation, "Microsoft Excel – speadsheet software – Office.com," 2013, URL http://office.microsoft.com/en-us/excel/

The MathWorks, Inc., "MATLAB – The Language of Technical Computing," 2013, URL http://www.mathworks.com/products/matlab/

1. REPORT DATE (DD-MM-YY) December 2014	2. REPORT TYPE Final	3. DATES COVERED (FROM – TO) November 2012 – January 2013		
4. TITLE AND SUBTITLE Discrete Reliability Projection		5A. CONTRACT NO. DASW01 04 0003		
		5B. GRANT NO.		
		5C. PROGRAM ELEMENT NO(S).		
6. AUTHOR(S) Snyder, J. A.; Fregeau, J. M.; D. Spalding		5D. PROJECT NO.		
		5E. TASK NO. AX-1-3100.42		
		5F. WORK UNIT NO.		
7. PERFORMING ORGANIZATION NAME(S) AND ADDRESS(ES) Institute for Defense Analyses 4850 Mark Center Drive Alexandria, VA 22311-1882		8. PERFORMING ORGANIZATION REPORT NO. IDA Document D-4808		
9. SPONSORING / MONITORING AGENCY NAME(S) AND ADDRESS(ES) Dr. J. Brian Hall Suite 07J22 4800 Mark Center Drive Alexandria, VA 22311		10. SPONSOR'S / MONITOR'S ACRONYM(S) DASD(DT&E)		
		11. SPONSOR'S / MONITOR'S REPORT NO(S).		
12. DISTRIBUTION / AVAILABILITY STATEMENT Approved for public release distribution is unlimited				
13. SUPPLEMENTARY NOTES David A. Sparrow, Project Leader				

14. ABSTRACT

We describe a model for reliability, growth projection against a discrete variable. Most reliability models only work for continuous variables. We include examples of this model's employment. One example demonstrates how a test event where the configuration was changed during the event can be reconciled. We also include Matlab and Excel code for applications of the model.

15. SUBJECT TERMS

Reliability, Reliability Growth, Discrete Reliability Growth Projection

16. SECURITY CLASSIFICATION OF:			17. LIMITATION OF ABSTRACT Unlimited	18. NO. OF PAGES 32	19A. NAME OF RESPONSIBLE PERSON Dr. J. Brian Hall
A. REPORT UNCLASSIFIED	B. ABSTRACT UNCLASSIFIED	C. THIS PAGE UNCLASSIFIED			19B. TELEPHONE NUMBER (INCLUDE AREA CODE) Phone: 571-372-2678